Machine Learning and Artificial Intelligence
A.J. Tallón-Ballesteros and C. Chen (Eds.)

doi:10.3233/FAIA200773

Deep Robot-Human Interaction with Facial Emotion Recognition Using Gated Recurrent Units & Robotic Process Automation

Suchitra SAXENA[1] , Shikha TRIPATHI and Sudarshan TSB
Faculty of Engineering, PES University, Bangalore, India

Abstract. This research work proposes a Facial Emotion Recognition (FER) system using deep learning algorithm Gated Recurrent Units (GRUs) and Robotic Process Automation (RPA) for real time robotic applications. GRUs have been used in the proposed architecture to reduce training time and to capture temporal information. Most work reported in literature uses Convolution Neural Networks (CNN), Hybrid architecture of CNN with Long Short Term Memory (LSTM) and GRUs. In this work, GRUs are used for feature extraction from raw images and dense layers are used for classification. The performance of CNN, GRUs and LSTM are compared in the context of facial emotion recognition. The proposed FER system is implemented on Raspberry pi3 B+ and on Robotic Process Automation (RPA) using UiPath RPA tool for robot human interaction achieving 94.66% average accuracy in real time.

Keywords. Facial Emotion Recognition, Deep learning; CNN, GRUs, Raspberry Pi II, Robotic Process Automation platform

1. Introduction

One of the prominent robotics research areas is the design of intelligent robots that can communicate and act as a human companion. With rapid advancements in hardware, computer graphics, robotic technology and artificial intelligence, more and more cobots and social robots have been designed. For a robot to act as a human companion, it should have emotional intelligence for effective human interaction. Emotions include cognitive evaluation, body language, expressions and feelings [1]. Emotion recognition is fascinating and a challenging task. Text, speech, facial expressions, biological signals and gestures can be used to recognize human emotions. Facial expressions have significant importance in the identification of human emotions in direct human to human interaction [2]. Researchers and engineers have attempted to design artificial intelligence frameworks, which are cognitively and/or physically similar to human behavior. The increase in computational power since a decade has largely contributed to the development of fast learning machines. In addition, the internet has generated a substantial amount of training data. These two development triggered research into smart

[1] Corresponding Author: Suchitra Saxena, Faculty of Engineering PES University, Bangalore, India; E-mail: suchitra@pes.edu.in, suchitrasaxena10@gmail.com

self-learning systems, with one of the most successful emerging techniques being deep learning networks. Progress in robot-human interaction over the past decade has contributed to several applications in robotic technology where robots need to comprehend human actions and emotions. With emotional intelligence robots, human action can be better predicted and performance can be increased in many applications such as Human Robot Interaction (HRI) for kid's therapy with autism and attention deficit hyperactivity disorder (ADHD), driver awareness alerting system, legal disciples, medical guides, E-Learning, psychology and entertainment feedback system, emotional support as socially assistive robots for kids [3-8]. Developing an algorithm, which can identify emotions from facial images, is therefore desirable to improve HRI. The advancement of automated processes, tools such as the Robotic Process Automation (RPA) have been successful in many fields in improving operational accuracy and performance. RPA is the implementation of automated robots or bots that use artificial intelligence techniques to automate repetitive tasks in real time. Also, implementation of the proposed architecture on the RPA platform would be relevant for HRI or Industry 4.0 applications [9].

In this work, a robust and efficient GRUs Facial Emotion Recognition System (GFERS) is proposed for recognizing emotions from facial expressions. In real time, the proposed method achieves 94.66% accuracy. CMU Multi-PIE [10] and FER 2013 databases [11] are used for training. GFERS is developed using GRUs, which is found to be efficient and robust in recognizing emotions under constraints such as head pose, illumination variations and age differences. The proposed system is deployed on Raspberry Pi3 B+ [12] and also implemented on UiPath robotic process automation tool which can be used in HRI applications. The performance of the proposed system is also compared with CNN and LSTM based facial emotion recognition.

The paper overview is as follows: In Section II related work and contributions are discussed, Section III describes proposed architecture of GFERS, Section IV explains results and analysis. Conclusion and future directions are discussed in Section V.

2. Related Work

Since several years, considerable work on recognition of emotions for HRI applications has been reported. Deep learning techniques, Convolution Neural Networks (CNN) and Recurrent Neural Networks (RNN) as applied to facial emotion recognition have been explored in recent years. In 2014, Jung H, et al., used CNN to recognize emotion in real time. They have used three convolution layers and two fully connected (FC) layers. They have used CK+ and FER-2013 databases for training and achieved 72.78% & 86.54% respectively [13]. In 2017, Chen et al. also used CNN with four convolution layers and two FC layers. They have used CK+ and JAFFE databases [14] and achieved 98.15% recognition accuracy for frontal pose. In [15], Saxena et al. proposed an algorithm based on CNN with three convolution layers and two FC layers with batch normalization to address overfitting. They have used Softmax classifiers and CMU MultiPIE databases for training. They have achieved average accuracy of 95.8% with pose and illumination variations in real time with 25 subjects. In [16], Baddar et al. used LSTM to address the influence of mode variability on the encoded spatio-temporal attributes. They have shown that LSTM encoded spatio-temporal features and retains a bias due to different variations such as illumination variation or pose variations by using static sequences. They have used Oulu-CASIA, AFEW and KAISI facial expression datasets and achieved

85.185, 51.44% and 84.98% accuracy respectively for these databases. In [17], Hasani et al. used an Inception-ResNet 3D CNN with LSTM to extract the spatial-spectral relationships within images and between different frames in the videos. They have used CK+, MMI, FERA, DISFA and achieved 67.52%, 54.76%, 41.93% and 40.51% respectively for head pose variations. In [18], Kim et al. used a spatio-temporal representation of a hybrid combination of CNN and LSTM to recognize facial expression and achieved 78.61% and 60.98% for MMI, CASME II databases respectively. In 2018, Yan et al. proposed a framework Joint Convolutional Bidirectional LSTM (JCBLSTM) to jointly model discriminative facial textures and spatial relationships between different regions [19]. They achieved average accuracy of 90.89% and 71.99% for different pose and illumination variations using Multi-PIE, FER-2013 databases. In 2019, Ilyas et al. proposed a hybrid CNN and LSTM model to address complexities and limitations of Traumatic Brain Injured (TBI) human-robot interaction. They used TBI-patient database, which is a collection of multimodal data annotated by physiotherapists, caretakers, experts, and doctors [20]. They have used CK+ database and achieved accuracy of 86.16% for pose and illumination variations. In [21], Deng et al. used 3D CNN algorithm for FER in videos. They used the framework of 3D Inception-ResNets structure, Stem layer, RNN special type GRU layer, Island layer, Dropout layer and Softmax layer to capture spatial relationships in facial expression images and temporal relationships between different facial frames. They have achieved average accuracies of 68.73%, 58.76% and 43.56% for CK+, MMI, AFEW databases. The proposed system is pose and illumination invariant. In [22], Li et al. proposed an HRI emotion recognition system. They first used CNN model to extract features from static images and later LSTM to find the relationship between the transformation of facial image sequences. They used CK+ database and achieved 90.51% accuracy for the frontal face. In 2019, Kang et al. proposed a CNN-RNN hybrid model. In which, first they used VGG16 to extract features from video frames and then used convolutional GRU to decode motion features [23]. They achieved 47% accuracy with the AFEW database for pose and illumination variations. In 2020, An et al. used CNN-LSTM hybrid model for feature extraction and Support Vector Machine (SVM) for classification [24]. They achieved average accuracy for pose and illumination variation of 98.9%, 99.3%, 86.6%, 87.7% and 88.3% for CK+, JAFFE, FER-2013, BU-3DFE, Oulu-CASIA respectively.

Most of the work reported in literature uses CNN and a hybrid of CNN-RNN. Mostly LSTM and GRUs are used for text classification and speech translation or classification tasks where sequence of data are used. To reduce training time and to capture temporal information LSTM and GRU can be used. Limited work is reported on LSTM based facial expression recognition, whereas use of GRUs based RNN for facial emotion recognition tasks will be more efficient. In this work, a facial emotion recognition system using RNN with GRU is proposed for facial images. We believe that using GRU for facial emotion recognition would reduce training time as compared to CNN and LSTM. The design of proposed architecture for the framework is described in the next section.

3. Proposed Architecture of GFERS

In Facial emotion recognition there are three main steps: first step is face detection, second is feature extraction and third is emotion classification. In the proposed work, face detection is obtained by using Viola-Jones face detection technique [25] and facial emotion recognition is implemented using Gated Recurrent Units (GRU) deep learning

technique. In GRUs model algorithm, GRU is used for feature extraction and followed by two dense layers for emotion classification as shown in Figure. 1. Layers details are given in Table 1.

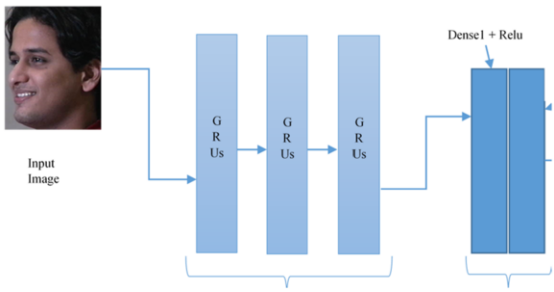


Figure 1. GFERS architecture

Figure 2. Few samples of Datasets used in GFERS

Table 1. Details of GFERS Layers

Input Image: 48x48x1	
GRUs Block	*GRU (512, Input_shape = (32,48,48), return_sequences = True)
	GRU (256, return_sequences=True)
	GRU (128, return_sequences=True)
	GRU (64, return_sequences=False)
Dense Block	Flatten: **FC1: 64 Activation function layer (ReLu)
	FC2: 6 Softmax

Notes: *GRU: Gated Recurrent Units Layer , **FC: Fully Connected layer

CMU MultiPIE and FER-2013 databases are used to train the layered model with six basic emotions (Happy, Anger, Neutral, Disgust, Sad and Surprise). The CMU MultiPIE database consists of 750,000 images of 5 basic emotions (Happy, Anger, Neutral, Disgust, and Surprise) posed by 337 subjects in different sessions. To add one more basic emotion (sad), we have used the FER-2013 database. FER-2013 database consists of 35685 images of 48x48 grayscale images with basic emotions (Happy, Sad, Neutral, Anger, Surprise, Anger and Disgust). The sample images used in training and validations are shown in Figure. 2. The algorithm of GFERS is shown in Table 2 and explained in the following section.

Table 2. GFERS Algorithm

Step1: Load training labeled data as
$S = S^{(1)}, S^{(2)}, S^{(3)}..,S^{(N)}$; N is the total class
Step2: Define Layers to extract and learn features
Input image Layer (Dimensions)
GRU: {*Units, Activation, Input_shape, return_seq}*
GRU: {*Units, Activation,return_seq}*
GRU: {*Units, Activation}*
...
Dense: {*Dimensional_vector*}
Activation function layer
Dense: {*Dimensional_vector =N*}
Softmax layer
Step 3: Training process
Set training options for training network Initial Weights $W = W_0$; Initial bias $\theta = \theta_0$
options = Training_Options_Adamwithproperties:
{Initial_Learning_Rate, beta_1, beta_2,Decay,MaxEpoch, MiniBatchSize}
[network, info] = train_Network (labelled_data, layers, options)
Step4: Testing Process in real time
Capture Image (webcam)
while (true)
Face = detect_faces (Haar_Cascaded classifier)
for face: Faces
Predicted Class = classify (net, Face)
classified _emotions (Display)
end
end

3.1. Feature Extraction

Gated Recurrent Units (GRU) layer is used for extracting features from images. GRUs are a new generation of Recurrent Neural Networks (RNN), which are a new gated mechanism introduced in 2014. GRU is equivalent to other RNN mechanisms like LSTM but has shown better performance on smaller datasets. GRU has only two gates; a reset gate and an upgrade gate, they omit an output gate as shown in Figure. 3.
GRU's use the hidden state to pass information. The function of an updated gate decides the set of data that can be retained for inclusion and remaining data to be excluded. The reset gate decides about how much of prior information should retain and discard. Equations for gated unit of GRU are shown below Eq. (1)-(3):

$$z_t = \sigma_g(W_z x_t + U_z h_{t-1} + b_z) \quad (1)$$

$$r_t = \sigma_g(W_r x_t + U_r h_{t-1} + b_r) \quad (2)$$

$$h_t = \{z_t \odot h_{t-1} + (1 - z_{t-1}) \odot \emptyset_h(W_h x_t + U_h (r_t \odot h_{t-1}) + b_h)\} \quad (3)$$

where,x_t, h_t, z_t and r_t are input, output, updated gate and reset gate vectors respectively. $W, and U$ are parameter matrices and b vector respectively. $\emptyset_h$ and σ_g are hyperbolic tangent and sigmoid activation functions respectively.Initially, for $t = 0$, the output vector $h_0 = 0$.

In this work, after repeated experimentation with different layers of combinations and units, it was found that better training and validation accuracy could be obtained using the proposed layered architecture. In this architecture there are four GRU layers with the first layer of 512 units, second layer of 256 units, third layer of 128 units and fourth with 64 units with input image dimensions of 48x48x1. In the preprocessing stage CMU MultiPIE database is resized from 205x260x3 to 48x48x1 for five emotions

(Happy, Anger, Neutral, Disgust and Surprise) and for sad emotion FER-2013 database is used without any modification.

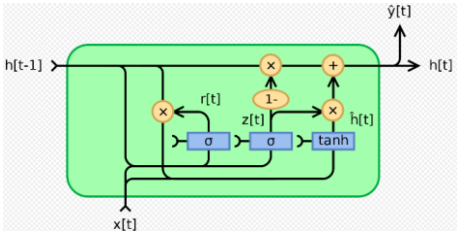


Figure 3. Gated Recurrent Unit

3.2. Classification

The GRUs output is fed to first dense layer as input for classification of emotion. The dimensional vector N = 256 with ReLu activation function is used for first dense layer output. The first dense layer output is given to second dense layer with N = 6 dimensional vector which is followed by Softmax classifier to get 6 emotion classes probability distribution using following Eq. (4):

$$S(p)_i = \frac{e^{p_i}}{\sum_{j=1}^{N} e^{p_j}} \quad (4)$$

where, p_i is the input vector, and since there are six emotions,p_j varies from 1,2,3…N=6.

4. Results and Analysis

In this section, GFERS performance evaluation and analysis has been discussed. 25 subjects are used for real-time performance testing, that resulted in an average accuracy of 94.66%. 2266 images per emotion were used (Happy, Neutral, Anger, Sad, Disgust, and Surprise), resulting in a total of 13596 images. The system's robustness and efficacy is measured under varying head pose and illumination conditions. It can recognize facial emotions from a distance between 0.30 m to 3 m approximately. The algorithm is implemented using Keras with Tensorflow as backend. GFERS is also implemented on RPA platform to make it easier to use in future robotic applications. The network is trained with adaptive moment estimation (Adam) with learning rate of 0.00008, a decay of 10-6 and beta1 and beta2 of 0.9 and 0.999 respectively for 359 epochs, batch size 32 and steps per epoch as 350 as shown in Table 3. The training accuracy and validation accuracy achieved by the model is 99.5% and 95.5% respectively. In real-time, the highest recognized probability value for each emotion is at-least 99%. GFERS achieved average recognition accuracy of 94.66% with 25 subjects in real-time as shown in Table 4. Single face and multi-face snapshots of results are shown in Figure 4 and 5 respectively under different illumination effect. GFERS is robust in various conditions of head pose and illumination and achieved good results in real time as shown in Figure 6. The system supported maximum head pose variation from frontal pose is ± 75º in yaw. It is also observed that GFERS could not detect multiple faces in few cases. In some cases, neutral is recognized as disgust, sad and vice versa [16]. The results are shown in Figure 7. Also, the proposed system is implemented for RPA platform using UiPath RPA tool [26], which is considered one of the industry's fastest and robust solution for RPA

implementation, since it helps the robots to make certain real-time process changes depending on the requirements of the tasks. The snapshot for RPA implementation is shown in Figure 8.

Table 3. Hyper parameters for proposed system GFERS.

Hyper parameter	Values
Adam optimizer	
learning rate	0.00008
decay	10-6
beta1	0.9
beta2	0.999
Epoch	359
Batch size	32
Step size per epoch	350

Table 4. GFERS Recognition rate

Emotion	Recognition Rate (25 Subjects)
Anger	92%
Disgust	92%
Happy	96%
Neutral	96%
Sad	96%
Surprise	96%
Average Probability	94.66%

Figure 4. Results of GFERS for single face under different illumination and head pose

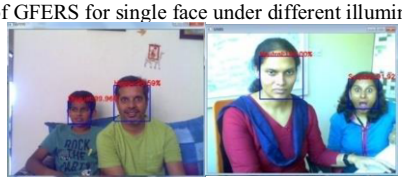


Figure 5. Results of GFERS for Multiple faces under different illumination and head pose

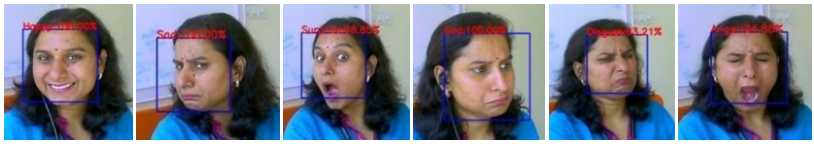


Figure 6. Results of GFERS for Pose (Maximum variation ±75º from frontal pose in yaw)

Figure 7. Failure results of GFERS

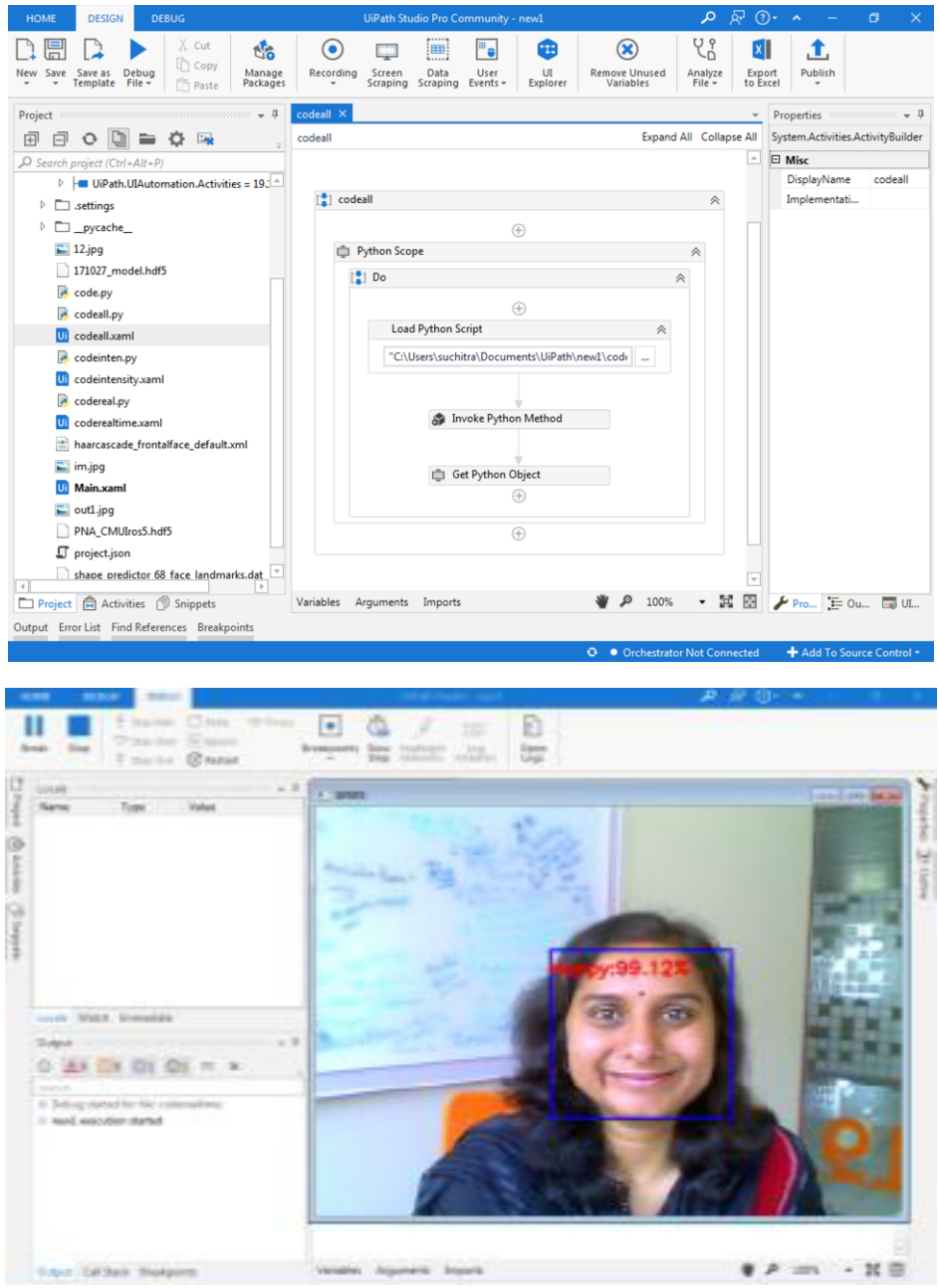


Figure 8. Snapshots of GFERS on UiPath RPA tool

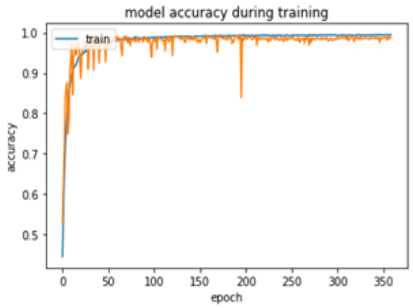
The GFERS results are compared with other related literature and a description of comparison is given in Table 5. The accuracy achieved by GFERS with various pose and illumination conditions is higher compared to the state-of-the-art techniques. In the existing literature, most of the reported work is based on CNN and CNN combined with

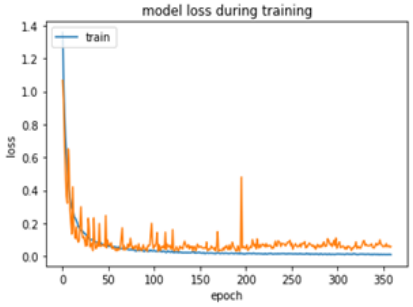
LSTM or GRU. Mostly LSTM and GRUs are used for text and speech classifications. For image classification, LSTM and GRUs are used for classification after feature extraction is done by using CNN. In the proposed architecture, standalone GRUs are used for feature extraction followed by dense and Softmax layers for classification achieving comparable accuracy with CNN and LSTM. Training time for GFERS, LSTM based model and CNN based model is 17.08 hours, 21.78 hours and 48 hours respectively. The training of all models is carried out by using same hardware platform (Intel i7-7700 CPU @ 3.60gHz and 16GbRAM). To train the model, use of GRUs took significantly lesser time compared to CNN models for the same number of epochs. The training time was reduced by a factor of approximately 30 hours making it feasible to carry out multiple experimentations with different combinations. The training process comparison of three model is as shown in Figure 9 (a)-(c).

Table 5. GFERS Results comparison with existing Literature (Real time)

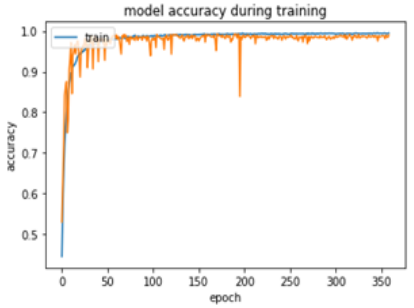
[Author, year]	Technique used	Data set used	Acc.*	PI*/II*/MF*
[Jung, 2014] [12]	DNN & CNN	CK, FER 2013	72.78% 86.54%	No/ No/No
[Chen, 2017] [13]	CNN	CK+, JAFFE	98.15%	No/ No/No
[Saxena, 2019][14]	CNN	CMU Multi PIE	95.8%	PI/II/MF
[Baddar, 2018][15]	Mode viriational LSTM	Oulu-CASIA, AFEW, KAISTFace MPMI	85.18% 51.44% 84.98%	PI/ II/No
[Hasani, 2017] [16]	3D CNN-LSTM	CK+, MMI, FERA, DISFA	67.52% 54.76% 41.93% 40.51%	PI/No/No
[Kim, 2019] [17]	CNN-LSTM	MMI, CASME II	78.61% 60.98%	No/No/No
[Yan, 2018][18]	Joint CNN-bidirectional LSTM	Multi-PIE, FER-2013	90.89% 71.99%	PI/II/No
[Ilyas, 2019] [19]	CNN-LSTM	CK+	86.16%	PI/II/No
[Deng, 2019] [20]	CNN-GRU	CK+, MMI, AFEW	68.73% 58.76% 43.56%	PI/II/No
[Li, 2019] [21]	CNN-LSTM	CK+	90.51%	No/No/No
[Kang, 2019] [22]	VGG16, GRU	AFEW	47%	PI/II/No
[An, 2020] [23]	Hybrid CNN-LSTM, SVM	CK+, JAFFE, FER-2013, BU-3DFE, Oulu-CASIA	98.9%, 99.3%, 86.6%, 87.7%, 88.3%	PI/II/No
GFERS	**GRU**	**CMU Multi PIE, FER-2013 (for Sad emotion)**	**94.66%**	**PI/II/MF**

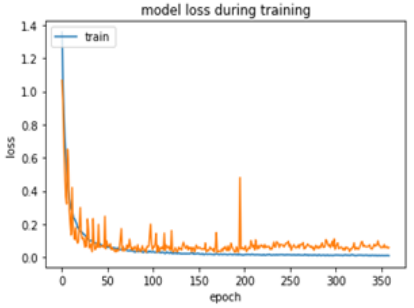
*Acc.: Average accuracy, PI: Pose Invariant, II: Illumination Invariant, MF: Multiple faces**NR: Not reported



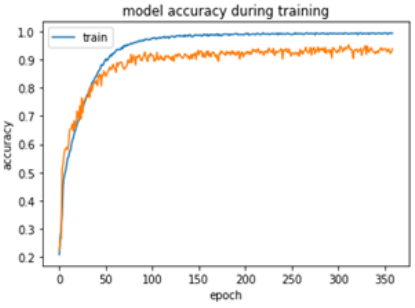


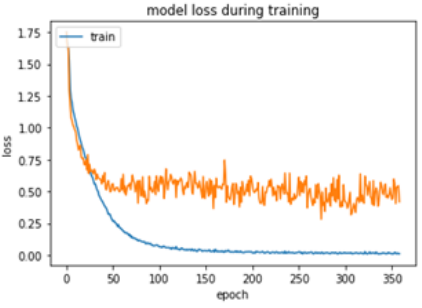
(a)





(b)





(c)

Figure 9. Comparision of Training Process for (a) CNN (b) LSTM (c) GFERS

5. Conclusion and Future Work

In this work, we have proposed a methodology for facial emotion recognition using GRUs. The proposed GFERS works successfully for multiple facial emotion recognition under the constraints such as pose, illumination and age variations. Average accuracy of 94.66% is achieved by proposed GFERS in real time for 25 subjects, which is higher than the results reported in existing literature. The probability of recognized emotion is almost 100% in some cases. In real-time input images, the processing time is in between 0.032-0.037 sec for CPU and 0.200-0.350 sec for Raspberry pi implementation respectively. The GFERS is also implemented for UiPath Robotic Process Automation making it suitable for HRI applications.

Future work involves developing algorithms to solve the constraints of occlusion and to test these techniques on social robots as a real time application.

6. Acknowledgment

The authors would like to thank all the volunteers for the experimentation and also would like to thank the host organization for providing CMU Multi PIE database. We thank all other researchers for making other relevant databases available for such research experiments.

References

[1] Dautenhahn K. Methodology & themes of human-robot interaction: A growing research field. Int. J. Adv. Robotic System. 2007; vol. 4, no. 1; p. 15.
[2] Mehrabian A. Communication without words. Psychology Today,2, 1968, pp. 53-56.
[3] Happy SL, et al. Automated alertness and emotion detection for empathic feedback during e-learning. IEEE 5th Int. Conference on Technology for Education (T4E); 2013; India; pp. 47-50.
[4] Coco MD, Leo M, Distante C, Palestra G. Automatic emotion recognition in robot-children interaction for ASD Treatment. IEEE Int. Conference on Computer Vision Workshop; 2015; Santiago; pp.537-545.
[5] Goodfellow IJ, et al. Challenges in representation learning: a report on three machine learning contests. Workshop Challenges in Representation Learning (ICM12013); 2013; pp. 1-8.
[6] Rosalind WP. Affective computing. MIT press, Cambridge. 2000.
[7] Suchitra, Palaniswamy S, Tripathi S. Real-time emotion recognition from facial images using raspberry Pi II. 3rd International Conference on Signal Processing and Integrated Networks, (SPIN), IEEE. 2016; Noida, India, pp. 666-670.
[8] Romao M, Costa J, Costa CJ, Robotic process automation: a case study in the banking industry. 14th Iberian Conference on Information Systems and Technologies, 2019, Portugal, pp1-6.
[9] Osman C, Ghiran A. When industry 4.0 meets process mining, Procedia Computer Science, 2019, Volume 159, pp. 2130-2136, ISSN1877-0509, https://doi.org/10.1016/j.procs.2019.09.386.
[10] Gross R, Matthews I, Cohn J, Kanade T, Baker S. Multi-PIE. Proc Int Conf Autom Face Gesture Recognit. 2010;28(5):807-813.
[11] Jung H, et al. Development of deep learning-based facial expression recognition system. 21st Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV). 2015; Mokpo; pp. 1-4.
[12] Goodfellow IJ, et al. Challenges in representation learning: A report on three machine learning contests. Neural Networks, 64:59--63, 2015.
[13] Warren Gay. 2014. Raspberry Pi Hardware Reference (1st. ed.). Apress, USA.
[14] Chen X, Yang X, Wang M Zou J. Convolution neural network for automatic facial expression recognition. International Conference on Applied System Innovation. 2017; Sapporo; pp. 814-817.
[15] Saxena S, Tripathi S, Sudarshan TSB. Deep dive into faces: pose & illumination invariant multi-face emotion recognition system. 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). 2019; Macau, China; pp. 1088-1093.
[16] Baddar WJ, Ro YM. Mode Variational LSTM robust to unseen modes of variation: application to facial expression recognition. AAAI (2018).

[17] Hasani B, Mahoor MH. Facial expression recognition using enhanced deep 3D convolutional neural networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 2017; Honolulu, HI; pp. 2278-2288.
[18] Kim DH, Baddar WJ, Jang J, Ro YM. Multi-objective based spatio-temporal feature representation learning robust to expression intensity variations for facial expression recognition. in IEEE Transactions on Affective Computing. 1 April-June 2019; vol. 10: no. 2; pp. 223-236.
[19] Yan J et al. A joint convolutional bidirectional LSTM framework for facial expression recognition. IEICE Transactions on Information and Systems. 2018;101(4); pp. 1217-1220.
[20] Ilyas CMA, Schmuck V, Haque MA, Nasrollahi K, Rehm M, Moeslund TB. Teaching pepper robot to recognize emotions of traumatic brain injured patients using deep neural networks. 28th IEEE International Conference on Robot and Human Interactive Communication. 2019; India; pp. 1-7.
[21] Deng L, Wang Q, Yuan D. Dynamic facial expression recognition based on deep learning.14th International Conference on Computer Science & Education (ICCSE). 2019; Toronto, Canada; pp. 32-37.
[22] Li TS, Kuo P, Tsai T, Luan P. CNN and LSTM based facial expression analysis model for a humanoid robot. in IEEE Access, 2019, 7; pp. 93998-94011.
[23] Kang K, Ma X. Convolutional gate recurrent unit for video facial expression recognition in the wild. Chinese Control Conference (CCC). 2019; Guangzhou, China; pp. 7623-7628.
[24] An F, Liu Z. Facial expression recognition algorithm based on parameter adaptive initialization of CNN and LSTM. The Visual Computer. 2020; 36; pp 483–498.
[25] Viola P, Jones MJ. Robust real-time face detection. Int. Journal of computer vision. 2004; 57,137-154.
[26] https://www.uipath.com/